C-V Test Structures for Metal Gate CMOS

Radko G. Bankras, Mark P. J. Tiggelman, M. Adi Negara, Guido T. Sasse and Jurriaan Schmitz

MESA+ Institute for Nanotechnology, Group Semiconductor Components
University of Twente, Enschede, The Netherlands
Tel. +31 53 489 2726 Fax +31 53 489 1034
1 Now at ASM International, Almere, The Netherlands
2 Now at Tyndall National Institute, Cork, Ireland

ABSTRACT
Gate leakage has complicated the layout and measurement of C-V test structures. In this paper the impact of metal gate introduction to C-V test structure design is discussed. The metal gate allows for wider-gate structures and for the application of n-p-n-p diffusion edges. We show, both theoretically and with experimental data, the impact of both design modifications on C-V measurement results.

INTRODUCTION
The gate leakage current density of next-generation CMOS technologies is forecast to extend beyond 1 kA/cm² [1], which complicates the characterization of MOS transistors with electrical techniques such as charge pumping and capacitance-voltage (C-V) measurements [2,3]. Increasing the measurement frequency to well above 1 MHz is one way to suppress the disturbing effect of gate leakage that gained considerable interest in recent years [4-8].

In this paper, an excursion is made to metal gate CMOS. Metal gates with their intrinsic benefit of much lower gate resistance offer a broader design margin for C-V test structures, as well as well-type implantations at the gate edge. We discuss the layout optimization of a high-frequency (or RF) C-V test structure given these new boundary conditions, and present first data from medium-k/metal-gate RF-CV measurements.

LOW SERIES RESISTANCE DESIGN
When going to higher frequencies, considerably more attention must be paid to the combination of measurement equipment, cabling, and test structures [7]. In test structure design, the two key issues to address are device series resistance on one hand, and non-quasistatic effects in the channel (in inversion) on the other. The non-quasistatic effects limit the channel length \( L \) roughly to \( \leq 1 \mu m \) for measurements up to a few GHz, see e.g. [9]. Designing for low series resistance is less straightforward.

In this section we summarize equations that can be used to estimate the external resistance of a MOS capacitor with transistor topography, i.e. a gate finger with width \( W \) and length \( L \), and with a highly doped source of minority carriers on two sides. We further connect the gate on both sides, and use an NMOS device as example. See figure 1. The connection of gate, source, drain and well are discussed in many previous publications (see e.g. [4,7]).

![Fig. 1. Test structure layout sketch for an NMOS capacitor with n+ diffusion edges and closely spaced well contacts. The key layout parameters are L, W and Ls. To reach the required overall capacitance, several of these structures can be drawn in parallel. The gate resistance of such a structure with silicided polysilicon gates was shown to be 10](image)

Referring to ref. [10] for further details, this equation expresses how the gate resistance is built up of gate sheet resistance \( \rho_{\text{g-sheet}} \), an additional term for the gate material outside the active region, the resistance in contacts between poly and metal-1, and the silicide-polysilicon contact resistance. For the clarity of reasoning, we simplify the above equation to

\[
R_{\text{gate}} = \frac{1}{2} \frac{W}{L} \rho_{\text{g-sheet}} + \frac{1}{2} \frac{W}{L} \rho_{\text{g-sheet}} + \frac{1}{2} \frac{R_{\text{via}}}{N_{\text{via}}} \frac{1}{W} \rho_{\text{con}} \frac{1}{L} \rho_{\text{con}}
\]

(1)

The gate resistance of such a structure with silicided polysilicon gates was shown to be [10]

\[
R_{\text{gate}} \approx \frac{1}{12} \frac{W}{L} \rho_{\text{g-sheet}}
\]

(2)

Approximation (2) holds well when \( W \) is large. (The overall dc resistance of a polysilicon line, as measured between the two contacts, is of course equal to \( W/L \) * \( \rho_{\text{g-sheet}} \). The factor \( 1/12 \) appearing in (2) is caused by two effects that reduce the small-signal resistance between the gate connection and the channel. Because the...
resistance between the contact and the transistor edge is considerably smaller than the resistance to the center of the gate finger, the distributed nature of this sheet resistance leads to an effective average resistance which is a factor 3 lower than the overall dc resistance [11]. By the connection of the gate on two sides, we gain an additional factor 4.) At the other side of the gate dielectric, the series resistance is either dominated by the sheet resistance of the inversion layer (in inversion) or by the well resistance (in depletion and accumulation).

In analogy with (2), the effective small-signal resistance of the inversion channel can be approximated as

\[ R_{\text{channel}} \approx \frac{1}{12} \frac{L}{W} \rho_{\text{ch,ch}} \]  

(3)

with

\[ \rho_{\text{ch,ch}} = \frac{1}{\mu_{\text{eff}} C_{\text{ox}} (V_{\text{gs}} - V_T)} \]  

(4)

A good \( C-V \) measurement in inversion is therefore obtained with a test structure where the design parameters \( W \) and \( L \) are chosen such that

\[ \frac{W}{L} = \sqrt[1/2]{\frac{\rho_{\text{ch,ch}}}{\rho_{n,g}}} \]  

(5)

at which point \( R_{\text{gate}} + R_{\text{channel}} \) shows a minimum. Since the channel resistance (~1 k\( \Omega \)) is much larger than gate resistance (~10 k\( \Omega \)) this calls for a long-width, short-channel device, as normally applied.

The well resistance is most relevant for accumulation and depletion capacitance measurements. When the well is contacted via the chuck, and the resistance of the wafer plays a role, an external resistance of a few hundred ohms can easily occur [12]. Top side contacts can reduce this resistance. A general expression for the well resistance is not transparent because of the freedom in topological solutions and the vertically varying doping concentration in the well [13]. Here, after [13] for simplicity we assume that the well contacts run parallel to the gate finger on both sides (as in figure 1), and also that the resistance below the source/drain regions is dominant. This leads to the expression

\[ R_{\text{well}} = \frac{1}{2} \frac{L}{W} \rho_{n,w} \]  

(6)

with \( \rho_{n,w} \) the sheet resistance of the well below a source/drain implantation (typically in the 0.1-10 k\( \Omega \)) range). To suppress the significance of \( R_{\text{well}} \), efforts should be focused on getting the well contact close to the channel, given a fixed technology (and hence fixed \( \rho_{n,w} \)).

The minimization of overall external resistance \( R \) can now be carried out using (2), (3) and (6) and given the sheet resistances in a certain technology. (It should be noted that the choice of \( W \) and \( L \) do not affect the relative significance of the oxide conductance \( g_{\text{ox}} \) and the oxide capacitance \( C_{\text{ox}} \) as both scale with gate area.)

**TEST STRUCTURES WITH METAL GATES**

From the above discussion it is clear that the layout optimization depends on the gate’s sheet resistance. Figure 2 shows how the quality factor in inversion for metal gates remains high when very wide transistors are laid out – while for poly gates the optimum quality factor is already reduced by gate resistance for gate fingers around 1 \( \mu \)m width. The trend for \( Q_{\text{opt}} \) is derived using (2), (3) and (6) and following the equation \( Q_{\text{opt}} = \{4 \ g_{\text{ox}} R (1+ \ g_{\text{ox}} R)/\}^{1/2} \) [4].

![Graph showing the relationship between gate width and quality factor Qopt for different gate types.](image)

**TABLE I**

VALUES USED IN THE MODELED CAPACITOR STRUCTURES

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value (inversion)</th>
<th>Value (accumulation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho_{n,g} )</td>
<td>0.03 or 10 k( \Omega )/( \square )</td>
<td>0.03 or 10 k( \Omega )/( \square )</td>
</tr>
<tr>
<td>( \rho_{n,ch} )</td>
<td>10 k( \Omega )/( \square )</td>
<td>n/a</td>
</tr>
<tr>
<td>( \rho_{n,ox} )</td>
<td>n/a</td>
<td>250 k( \Omega )/( \square )</td>
</tr>
<tr>
<td>( L )</td>
<td>0.1 or 1 ( \mu )m</td>
<td>0.1 or 1 ( \mu )m</td>
</tr>
<tr>
<td>( g_{\text{ox}} )</td>
<td>1 MS/m²</td>
<td>1 MS/m²</td>
</tr>
<tr>
<td>( C_{\text{ox}} )</td>
<td>15 mF/m²</td>
<td>15 mF/m²</td>
</tr>
</tbody>
</table>

The same trend is found in depletion and accumulation, because \( R_{\text{well}} \) and \( R_{\text{channel}} \) have the same \( W \) dependence ((3) and (6)). Summarizing, the switch to metal gates leads to a wider freedom in the choice of \( W \), yielding a higher optimum \( Q \) value in some cases.

The well resistance can be drastically reduced when a p+ and an n+ implantation are used on either side of the gate. In polysilicon gate technologies this leads to uncontrolled gate doping; however, a metal gate is not affected by the source/drain ion implantations. A structure as in figure 3 can now be fabricated, with a p+ edge and an n+ edge around each gate. The two-gate layout yields some compensation of the lithographic misalignment between the n+ and p+ implantation masks. Note that the effective channel resistance in
inversion will quadruple, because of the single-sided minority carrier supply for each gate!

With the p⁺ implant close to the gate, a smaller overall test structure area is needed for a given \( W \times L \) overall capacitor dimension. With larger \( W \) per gate finger, the structure is also more compact because less gate fingers are needed. A metal-gate \( C-V \) test structure can thus be laid out in a more compact manner than a polysilicon-gate test structure, especially when the multifinger approach is used.

![Fig. 3. Layout of a metal-gate MOS capacitor, with an n⁺ and a p⁺ diffusion edge around each gate, facilitating the supply of both electrons and holes. Top: top view. Bottom: cross-sectional view.](image)

**Experimental Results**

Test structures as depicted in figure 3 were manufactured with an experimental process in the MESA+ Clean Room. The process features 1.5 \( \mu \)m lithography, a 1.6-2.1 nm EOT Al₂O₃ gate dielectric, and a TiN/Al gate. A 5-10 \( \Omega \)cm p-type wafer was used with a p-well implant. The dielectric and TiN capping are ALD-deposited without vacuum break, in a home-built ALD cluster system. The technique results in a very thin interfacial SiO₂ of 0.3 nm. Note that the channel length can in principle be laid out in submicron dimensions in spite of the limited lithographic capabilities: the effective channel length is determined by the offset between n⁺ and p⁺ masks (and implantation straggle and lateral diffusion).

Two-port S-parameter measurements were carried out on these test structures between 50 MHz and 2 GHz, following SOLT calibration and open-short de-embedding routines as previously described e.g. in [4]). The capacitance was computed from \( Y_{gg} \) using a three-element equivalent circuit and employing a two-frequency technique [14, 15] (at 100 and 200 MHz) to correct for series resistance and gate leakage.

Figure 4 shows \( C-V \) curves as obtained from three devices with different gate dimensions. On several wafers with different Al₂O₃ thickness and dielectric leakage, similar curves were found. Increase of the capacitance in the inversion region (“inversion capacitance”) is only observed in a small subset of the measurements, when the frequency is low (\( \leq 100 \) MHz) and the drawn gate length is shortest (1 \( \mu \)m). We attribute this to the combination of high threshold voltage (> 1 V), reduced channel mobility (because of the medium-k dielectric with thin interfacial oxide) and the long channel which is only connected to a minority carrier source on one side (see figure 3). It is expected that inversion capacitance is normally observed on similarly designed devices with channel lengths below 0.5 \( \mu \)m.

![Fig. 4. C-V curves of three capacitor structures with various width and length. The parasitic capacitance leads to a capacitance offset of all curves. The accumulation and depletion behavior is reproduced; inversion is not apparent (see text).](image)

To derive a specific capacitance, structures with several n⁺-p⁺ spacings (‘channel lengths’) are required. It is essential to verify the proper length and width scaling of measured gate capacitance, because through this verification many instrumental problems can submerge. In figure 5 we show the width and length scaling of the manufactured test structures in accumulation. Proper scaling is observed and through subtraction of capacitance at different length and width, the intrinsic (channel) capacitance per unit area can be computed. A possible different approach is to directly subtract \( Y \)-parameters of devices with various gate length and gate width to obtain the intrinsic admittance of the channel region, followed by capacitance computation. For this approach to work well, \( g \) should scale properly with gate area.
The quality factor of the fabricated devices indeed shows a frequency dependence corresponding to the three-element model. This is illustrated by figure 6, where the quality factor as a function of frequency is displayed for a few devices. The decreasing $Q_{\text{opt}}$ with increasing $W$, as predicted in figure 2, is well visible. The deviations from the three-element model at higher frequencies, as observed by Jeamsaksiri et al. [8], were not found on these devices. A possible explanation is the lower gate resistance in our case, justifying the representation of the gate resistance with a single lumped element.

Fig. 6. Quality factor of three test structures with different gate width ($L = 1 \mu$m). The symbols represent measurement results; the lines represent 3-element model trends and are drawn to guide the eye.

**CONCLUSIONS**

In this paper the impact of metal gate introduction to $C$-$V$ test structure design is discussed. The metal gate allows for wider-gate structures, leading to a test structure with less overall area, and higher quality factors at the higher measurement frequencies. An extremely low series resistance in accumulation can be achieved with the application of $n$-$p$ diffusion edges (a gated diode configuration). However, this leads to a four times higher external resistance in inversion. The experimental results show the validity of the three-element approximation for the layouts, allowing for a straightforward capacitance extraction.

**ACKNOWLEDGEMENTS**

We wish to thank Tom Aarnink for technical support with the ALD processing, and Henk de Vries of the department of IC Design for his assistance with the RF measurements. The work of R. G. Bankras and G. T. Sasse was funded by the Dutch Technology Foundation STW.

**REFERENCES**